

Web-based Content Organization and the Transformation of Traditional Classification Systems

Joseph Busch, Taxonomy Strategies
Washington, D.C., USA

Abstract

Traditional hierarchical classification systems were designed to optimize the management, findability and use of physical collections of objects. But as more and more collections of objects are being accessed and used on the Web, these predominant classification models have been modified by faceted taxonomies with semantic relationships. The diverse uses of information require specialized classification strategies that reach beyond simple use cases. The ubiquity of web search engines and hypertext is also leading to new interest in labelling and describing named entities. This paper discusses these developments and provides examples for their utility in a variety of organizations, profit and non-profit.

Locating Objects

Traditional or global classification schemes respond to the need to physically locate objects in one dimension. In the classic example, a library book will be shelved in one and only one location, among an ordered set of other books. Thus the development and adoption of library classification systems including the Dewey Decimal Classification, Universal Decimal Classification, and Library of Congress Classification.¹ Traditional journal tables of contents similarly place each article in a given issue in a specific location among an ordered set of other articles, certainly a necessary constraint with paper journals and still useful online as a comfortable and familiar context for readers.

In the commercial realm, supermarkets and department stores with a large assortment of products have departments and sub-categories to assist location. They may vary from market to market, but the general schemes are common knowledge to most shoppers in a given region.

For example, for food markets in North America, we find categories like:

¹ These classifications systems have been the subject of a long and well-documented critique in the field of Library and Information Science. See for example, Bowker & Star (1999), Foskett (1971), Furner (2007), Furner & Dunbar (2004), Olson (2002), Olson & Schlegel (2001), and Tennis (2002).

- Bakery,
- Beverages and Snacks,
- Dairy,
- Deli and Prepared Foods,
- Frozen Foods,
- Grocery,
- Household,
- Meat and Seafood, and
- Produce.

And for department stores, categories like:

- Men's Clothing,
- Women's Clothing,
- Baby and Children's Clothing,
- Home Furnishings,
- Electronics,
- Toys and Sports, and
- Food.

Again, the problem solved by these classification schemes is to locate specific products in a primary location so that shoppers can readily find them.

Hierarchical Classification Problems

But in collapsing categories to one dimension, a traditional classification scheme makes essentially arbitrary choices that have the effect of placing some related items close together while leaving other related items very distant from each other. Continuing with the supermarket example, locations made in terms of storage (shelf, cooler, or freezer) may not reflect their ultimate use. Vegetables may be shelf-stable (preserved in jars or cans), fresh or frozen. These are essentially the same ingredient, but stored and merchandised in disparate locations. Similarly, tortillas are typically found in a U.S. supermarket in the dairy cooler, on the bread shelf and/or in the Hispanic food section. Brick and mortar retailers moving to online sales cannot simply replicate the store layout. The result has the effect of repeating the terms associated with the last dimension in many different contexts, leading to an appearance of significant redundancy and complexity in locating terms.

To illustrate this further, the classification of scientific literature can quickly become very complex. The *Physics and Astronomy Classification System* (PACS 2010) developed by the by the American Institute of Physics (AIP) for classifying scientific literature is a traditional classification system with a monolithic hierarchical set of codes. As shown in Figure 1, there are at least 62 different categories in PACS related to the term “semiconductor”. These occur in different contexts, primarily organized by broad physics disciplines such as “Materials Science” or “Condensed Matter”.

However, understanding the properties of semiconductors relies on quantum mechanics which is in the “General” PACS category. Condensed Matter itself is such a large discipline that it is split into two of the broadest PACS categories, and semiconductor-related categories occur in all three of these broad divisions. This repetition of “semiconductor” is an example of the redundancy that tends to occur in mono-hierarchical classification schemes. This makes these schemes difficult to navigate, and difficult to use, especially by those who are not information professionals. For example, authors of scientific articles that are being submitted to an American Physical Society (APS) journals are required to select the appropriate PACS classification as part of the article submission process. That selection in turn drives the selection of the appropriate APS journal (APS publishes 12 journals divided by Physics disciplines) and aids in the selection of the submission referee. APS also holds a major conference called the March Meeting which has more than 10,000 attendees, (as well as smaller meetings related to various Physics sub-disciplines). Grouping papers by topic so that sessions make sense, and also so that attendees can

physically get from one session to another is complicated. Conference planning is currently done by convening a large “sorting” meeting where papers are broadly grouped and then manually sorted by hundreds of volunteers.²

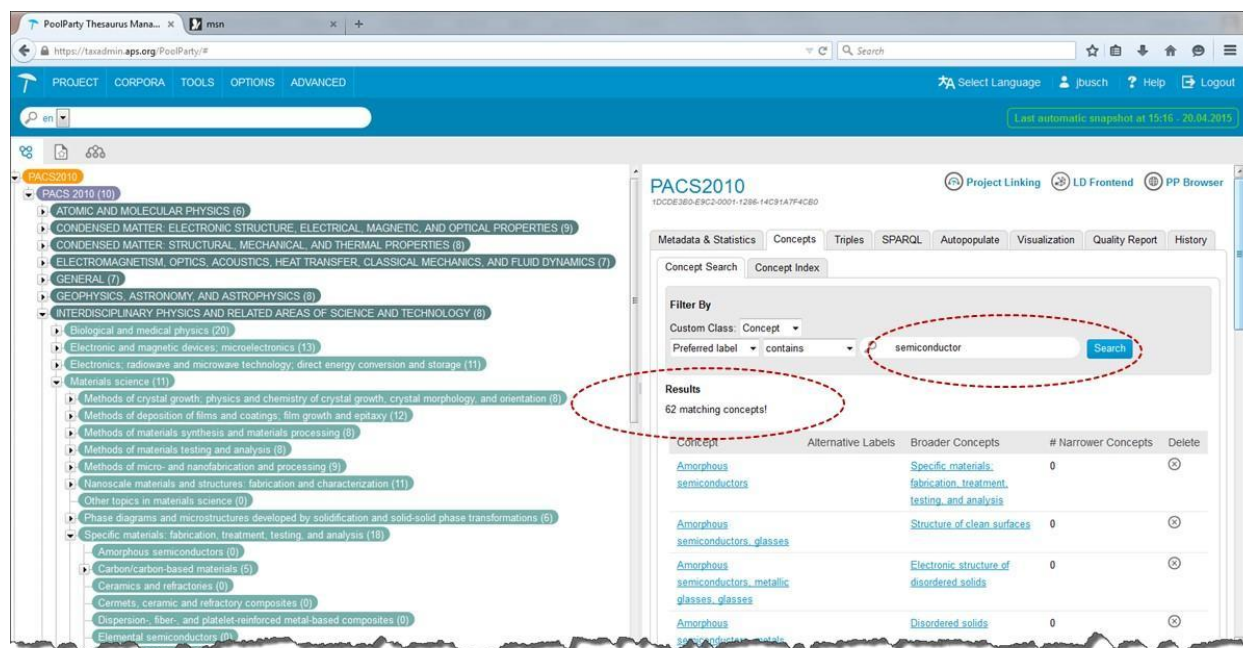


Figure 1-The term "semiconductor" occurs in 62 different PACS classifications.

Multidimensionality of the Real World

As shown in the above examples, the real world of things (products) and concepts is multi-dimensional. This is manifested in online shopping and elsewhere. It has become common to refine searches with filters on consumer product websites as well as content-based websites. Zappos (zappos.com), an online shoe and clothing retail business, uses the following attributes to filter a search on Men's Sandals which returns nearly 2,000 products:

² Note that AIP recently replaced PACS with the newly developed *AIP Thesaurus*; and in 2015, APS is introducing a new faceted classification scheme for physics that will replace the use of PACS in their scholarly publishing operation. I discuss this in more detail below. Changing to a new classification scheme presents “problems and opportunities” such as migrating to the new scheme without having to reclassify all the editors and referees, or maintaining backward compatibility or not. There are costs and benefits associated with the decision to change classification systems.

- Men's Size
- Men's Width
- Occasion
- Styles
- Color
- Brand
- Price
- Materials
- Insole
- Theme
- Pattern
- Accents

The Robert Wood Johnson Foundation (rwjf.org), a United States non-profit health policy philanthropy, uses the following attributes to refine searches on their content-based website:

- By Topic
- By Content Type
- By Age
- By Gender
- By Race/Ethnicity
- By Location
- By States and Territories

Filtering search results invites end users to refine their search results without having to type in a new search. It exposes contextually relevant metadata attributes and usually indicates how many matching “hits” will remain when the filter is selected. Sometimes it is also easy to remove a filter and select a different one. In these ways, faceted navigation allows a user to explore a collection of search results, drill down into those results by applying one or more filters, or remove a filter. This is an active use of multi-dimensional classification to help users explore a richly categorized collection of items.

While online shopping has become a commonly understood metaphor, applying multi-dimensional classification and faceted navigation to content collections is not so intuitive. Application users do not always recognize the purpose of or use the search filters in the right or left rail of a user interface. Sometimes users revert to the search box, expecting the type and go “I’m feeling lucky” Google experience, instead of a “shopping for shoes on Zappos” experience.

Complex Classification Use Cases

Clearly, the uses of classification systems in the real world are sometimes more diverse and complex than simply ordering a set of related content items in a search results set. That outcome is still important, but understanding how any given organization or individual might actually use – or wish to use – the organized information is critical. A “use case” explores these various scenarios with multiple stakeholders. Using formal and informal interviews, coupled with quantitative data, as well as learning about organizational goals and expectations, potential activities or likely uses can be developed for a given set of organized information. These “use cases” can be limited to the internal use, or they can include both internal and external activities. They facilitate the development of a specialized taxonomy that describes a variety of activities and uses, or contexts that are important for particular applications in particular settings.

Returning to the example of a scholarly publisher, the primary use case of a classification system for the American Physical Society is to facilitate an efficient and effective editorial and publishing process in order to be able to process tens of thousands of articles and papers each year. Organizing and facilitating

the editorial and publishing process at a scholarly publisher like the APS includes the following activities or use cases for of their classification system:

- Selection of taxonomy terms (indexing) for articles,
- Authors' assigning topics to their submissions,
- Defining areas of responsibility and interest for editors,
- Assigning articles to APS editors,
- Referees describing their areas of expertise,
- Selecting referees to review articles,
- Assigning articles to journal sections, and
- Generating statistical reports and lists of articles by various subject criteria.

These diverse uses of information require classification strategies that reach beyond those available in traditional classification systems.

For a multinational computer technology company like Dell, the primary use case is to facilitate the identification and linking of a large and changing collection of content items with a large and changing assortment of related products. However, this “big use case” is the overall strategy for locating products so consumers can buy them. It is also necessary to break this down into more specific tactics or steps that can be implemented in the user interface. Some of the specific tactics that Dell identified in 2013 to improve the effectiveness of their website were to:

- Improve organic (Google) search ranking by effectively incorporating synonyms in web content.
- Provide a consistent user experience across websites.³
- Use consistent navigation labels for products and services on the website and consistent terminology in the content.
- Use technology content to pivot between service and product.⁴
- Associate contextually relevant learning content with specific products and services.
- Provide links from learning content to specific product content.
- Provide contextually relevant navigation with industry solutions website content.
- Provide contextually relevant and consistent navigation among Dell solutions destinations (including Solutions, TechCenter and blogs) to share solutions content and best practices.
- Consolidate community content in a single user experience.
- Unify support and community content.⁵
- Integrate contextually relevant product support information with product details.
- Surface contextually relevant software and peripherals information.⁶
- Provide contextual navigation that highlights parts categories related to product category, and links to parts that are related to the specific product.
- Implement a method to tag content by segment so that global changes can be made to re-label, or merge segments (called content de-segmentation).

³ In 2013 Dell had separate websites for consumer, small and medium business, large enterprises, government, and education segments. Today, the Dell website merchandises to two segments – Home and Work. Dell has always sold to large organizations using direct sales representatives.

⁴ Dell was trying to change (“pivot”) from being primarily a product company to providing various types of IT services using, e.g., cloud computing platform rather than just selling consumers Dell servers and storage devices.

⁵ Since 2013 it has become a common practice to create user forums as a key component of user support.

⁶ Dell has an extensive software and peripherals catalog of products primarily from other vendors. This online store is currently called “Electronics & Accessories”.

- Provide contextual navigation when accessing external content.

This list breaks down the big use case “locating products so people can buy them” into a large number of tactical steps. Even so, the Dell use cases can be grouped by the type of information architecture methodology that should be used to address them. These are summarize in Table 1. Visiting the Dell website in 2015, one can notice that many of these use cases have been addressed over the past two years. Even though the sheer number of use cases implies complexity, the actual integration requirements break down into just a few patterns and best practices that can be widely applied across the online collection.

Use Case	Contextual Navigation	Site Architecture	Synonyms	Import Files
Improve Google search.				
Consistent experience across sites.				
Consistent terminology.				
Use technology content to pivot between service and product.				
Associate educational content with specific products.				
Move from educational to product content.				
Provide context within industry solutions.				
Consistent solutions and best practices.				
Consolidate community content.				
Unify support and community content.				
Integrate product support with product details.				
Surface software and peripherals information.				
Surface parts and accessories with products.				
De-segmentation.				
Integrate external content.				

Table 1-2013 Dell website performance improvement use cases

Importance of Facets and Relationships

These examples of real-world classification used by online shopping websites such as Zappos and Dell, and content websites such as Robert Wood Johnson Foundation and the American Physical Society (APS) illustrate how traditional classification systems now require new methods of content organization on the Web. The complex use cases discussed above are well-served by the classification methods of 1) facets and 2) semantic relationships. Facetted classifications deconstruct complex concepts into a grammar expressed as statements of named entities modified by types and topics. The key semantic relationships that are commonly manifested in web classifications are equivalent (synonyms), hierarchical (broader/narrower) and associative (related) relationships. Faceted classification and semantic relationships are important contributions that are actively transforming traditional classification systems as they are used on the Web, and for use with digital content repositories.

Facetted Taxonomy Examples

APS is in the process of implementing a new facetted taxonomy to replace the Physics and Astronomy Classification System (PACS) which has been discussed above (and has been used by both APS and AIP). APS submissions have required authors to identify the PACS code under which their submission should be categorized. That code has been subsequently used to assign the article to an APS editor, to select referees, and ultimately to assign the article to a category in the journal table of contents. The new taxonomy will replace PACS as the tool to facilitate the article submission, refereeing and publication process. One early idea for conceptualizing the new APS taxonomy broke down the description of physics research into the following components:

Research Description Component	Taxonomy Facet
• What you are studying	• Broad area, materials and systems
• Why you are studying it	• Phenomena and properties
• How you go about studying it	• Apparatus, theory and techniques

This method for description is easy to explain to researchers, and easy for them to learn. It breaks up a complex categorization task into smaller chunks. It is no longer necessary to parse large sections, or the whole hierarchical classification scheme to find the single most appropriate category. This is likely to result in more complete and consistent categorizations. More complete and consistent categorizations will create a collection that will also be easier and more effective to use to support various purposes.

Taxonomies are often developed to help organize commonly generated business information that exists in many forms and formats. These may be intranets, document management repositories or simply shared file directories with files and documents to support common business functions such as marketing and communications. Regardless of whether this is a commercial enterprise, government agency or NGO (non-governmental organization), common facets apply to all forms of organized information. These taxonomy facets include: Content Type, Audience, People, Organization, Industry, Location, Function, Product and Topic and are described in Table 2.

Facet	Definition	Example Source
Content Type	Types of content created, managed and used to record or communicate information.	AGLS Document Type (AGLS) , AAT Information Forms (AAT), Records management policy, etc.
Audience	Subset of constituents to whom a content item is directed or intended to be used.	Market segments, Educational stages/grade levels, etc.
People	Names of important people such as authors, politicians, leaders, actors, etc.	Library of Congress Name Authority File (LCNAF), NY Times Topics-People (NY Times), etc.
Organization	Names of organizations, their aliases and the relationships between them.	LCNAF, NY Times Topics-Organizations, etc.
Industry	Broad market categories such as industry sector codes.	North American Industry Classification System (NAICS), International Standard Industrial Classification (ISIC), etc.

Facet	Definition	Example Source
Location	Names of places of operations, activities, constituencies, etc.	Country Names (ISO 3166), Geonames (USGS), NY Times Topics-Places, postal services, etc.
Function	Activities and processes performed to accomplish goals.	Federal Enterprise Architecture Business Reference Model (OMB), AAT Functions, etc.
Product	Names of products and services that are produced by an organization or people.	Household Products Database (HHS), United Nations Standard Products and Services Code (UNSPSC), etc.
Topic	Topical subjects and themes that are not included in other facets.	Library of Congress Subject Headings (LCSH), NY Times Topics-Subjects, etc.

Table 2-Commonly used real-world taxonomy facets

Similar to the APS scholarly publishing example, a faceted taxonomy changes the categorization task from one where the problem is to find the best single place to file a content item (the goal of traditional classification systems), to one where the task is to describe the various attributes of a content in order to scope its context (the transformed goal of 21st century classification). Context is specified by describing multiple aspects of a content item – For a business item: What type is it? Who was it created for? What business activity is it related to? What people, organizations and/or products is it about? Is it related to particular location, industry sector or market? etc. Facetted classification is more like filling in the attributes of a product, than choosing the single most important aspect of a content item. Breaking up the categorization task into several discrete categorizations makes the process easier to accomplish. It is more often completed, and it is more consistently practiced.

Semantic Relationships

Before the World Wide Web, online searching was primarily limited to expensive abstracting and indexing information services. Today most people use free or inexpensive web search services. These so-called “organic” search engines (Google, Bing, Yandex, etc.) have become ubiquitous, meaning that everyone uses them all the time. Web search engine results have been optimized using analytics based on 1) co-citations (what is linked what), 2) keywords (what strings retrieve what pages), 3) popularity (what pages do most people view), and 4) any other relevance predicting factors that emerge that can be observed. But recently there has been interest in semantic methods to improve web search engine and website search information retrieval.

The key semantic relationships that can be used to improve information retrieval in search engines are equivalent, hierarchical and associative relationships. These relationships are also the basic components of the model for expressing classification schemes, taxonomies, thesauri and other similar knowledge organization schemes. Equivalent terms are alternative labels such as a synonyms or quasi-synonyms⁷.

⁷ Quasi-synonyms are terms that are not normally synonyms but are deliberately made into synonyms for a specific reason, e.g., rolling up narrower concepts into a broader concept for information retrieval purposes.

Hierarchical terms are broader (more general) or narrower (more specific) concepts. Associative terms are concepts that are inherently related but are not more general or more specific than each other.

Identifying, recording, managing and using equivalent terms can be very helpful in improving search engine performance. The difficulty is in determining how to appropriately apply the concept of equivalence. An “exact match” is a very high bar to apply, yet in some applications such as medical diagnostic applications, an exact match is critical. On the other hand matching codes to descriptive labels, or acronyms and abbreviations to spelled-out labels can be very helpful in improving search results. On the other extreme, in business applications it may be quite effective to roll-up narrower concepts into a broader category, for example, when there is not a critical amount of content in those narrower categories. Name authority files such as the Getty Trust’s *Union List of Artist Names* (ULAN) are very helpful for search applications that are federated across multiple institutions, or for web searching where there is little control over the way that names are spelled. Figure 1 is an excerpt for the ULAN record for the famous Dutch painter Rembrandt van Rijn. This record include 31 vernacular versions of the artist’s name.⁸

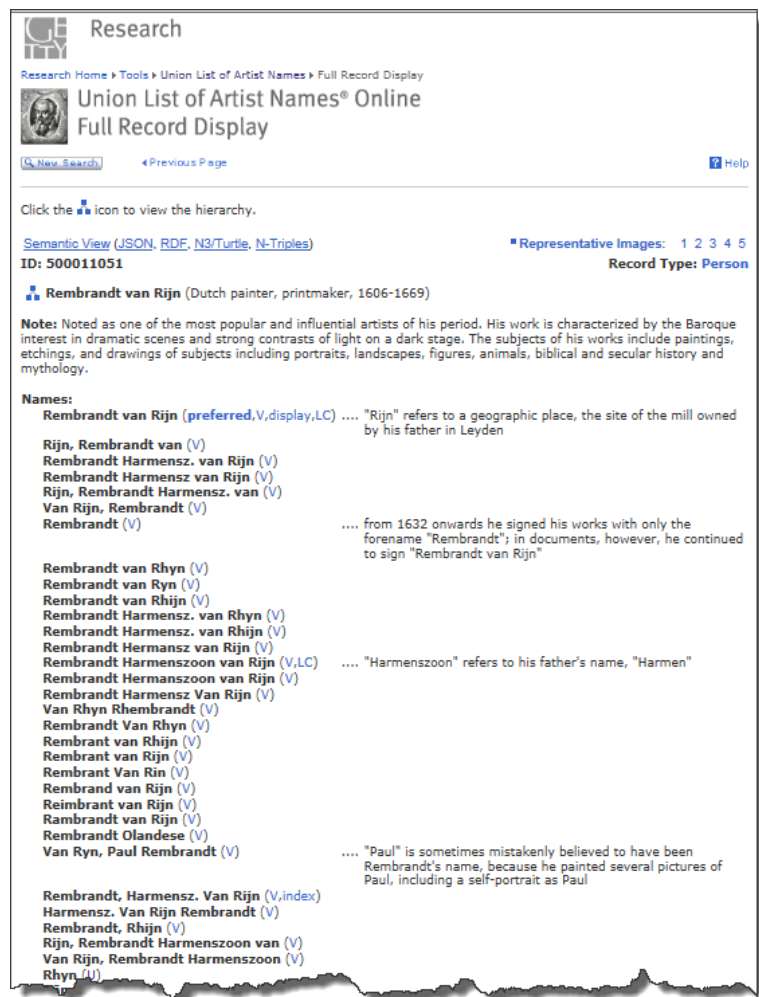


Figure 2-Getty Union List of Artist Names entry for Rembrandt

⁸ It should be noted that owners of a Rembrandt can probably spell the artist’s name any way they want.

Hierarchy in term lists is a very common way to group sets of related terms together. The two common types of hierarchy are 1) whole-part hierarchies (such as geographic locations), and 2) hierarchies which consist of instances (such as classified lists). It's also possible that the same terms may be members of more than one group. A common example is a geographic taxonomy where Greenland is listed both as part of North America as well as part of Denmark. Corn is a commodity as well as a vegetable.

Related terms are links made between concepts that are not hierarchical. Information scientists require associative relationships to have a critical mass of semantic or conceptual association to qualify for making an explicit linkage. The guidelines can quickly expand, causing difficulty for interpretation and use.⁹ But one of the most basic architectures of the World Wide Web has been hypertext, or making explicit links between things. The Google Knowledge Graph (Google) is an effort to curate hyperlinks, where the relationships between entities are used to present related content. For example, the Google search on "Rembrandt" returns search results plus a Knowledge Graph in the right rail. This includes images (self-portraits are displayed); a snippet from Rembrandt's Wikipedia entry; birth and death dates and locations (locations are hyperlinked to knowledge graphs for the locations); nationality (Dutch); a list of artists who Rembrandt influenced (Vincent van Gogh, Gerrit Dou, Jan Lievens, Carel Fabritius, Govert Flinck); periods of his work (Baroque, Dutch Golden Age painting, Dutch Golden Age) ; works (images with dates); plus related searches (what Google keeps track of). Knowledge Graphs are automatically generated by Google based on their analysis of the World Wide Web, but in some cases (for named entities like companies and living people with a logon and passcode) can be edited and annotated.

Another trend on the Web is the specification of schemas to describe and mark-up information about named entities. Schema.org, sponsored by the leading web search engines, provides schemas for creative works (books, movies, TV shows, music, recipes, etc.), events, organizations, persons, places, products, and other entities. The goal is to embed more machine-readable mark-up that identifies named entities and the relationships between them in web published content. Search engines are already incorporating rich descriptions of named entities when appropriate on in search results pages. For example, the Google search on American Art Museum displays a rich description of this named entity which is a part of the Smithsonian Institution in the right rail. This description shows other parts of the Smithsonian Institution such as the Hirshhorn Museum, art works that are owned by or on exhibit at the museum such as the fifty-one channel video installation "Electronic Superhighway: Continental U.S., Alaska, Hawaii" by the artist Nam June Paik, and related information such as glass canopy designed by Norman Foster.

We are surrounded by named entities and clues about the relationships between and among them. While social network analysis has been used by historians for a long time, the experience of the World Wide Web has made this transparent to everyone who participates in the web. The challenge of making sense of all this "stuff" has made information retrieval the key enabling technology for the Web. Underpinning web search is the analysis and annotation of the relationships between entities. This is a continuous feedback mechanism that attempts to improve search results and enhance the presentation of relevant related information.

⁹ See, e.g., *NISO Guidelines for the Construction, Format, and Management of Monolingual Thesauri*. ANSI/NISO Z39.19 – 2003. "Section 5. The Associative Relationship", pp. 19-20.

Conclusion

The Schema.org emphasis on identifying named entities ultimately ends with unique identifiers or URI's (Universal Resource Identifiers such as URL's) for named entities. Interestingly, this is not so different from traditional classifications whose codes aim to provide unique locations for physical objects. In this way perhaps we have come full circle from universal classifications systems to systems of universal unique named identifiers.

While the stakeholders and infrastructure aren't quite the way they were envisioned at the Getty's Art History Information Program back in the 1990's, the ecosystem for distributed vocabulary development actually exists today. We imagined a distributed environment in which many various types of stakeholders would contribute their vocabularies to common resources for creators (ULAN), geography (TGN) and material culture (AAT). Contributors would be able to get back enriched semantic resources and services (what we would call a web service today). One of the key missing ingredients was a system of persistent unique identifiers that could be assigned to semantic concepts. Now that we have this system of identifiers will semantic consortia emerge to collaborate on domain-specific vocabulary systems and services?

The open data movement has made it easier to publish semantic resources, but the hard work of building them still remains. Schema.org and Knowledge Graphs provide one path for building semantic resources. It is important to participate in these ecosystems even though they often lack the level of quality and control that might be desired. I believe there is more awareness and interest in semantic resources in the private sector than we might imagine. The challenge is how to identify opportunities and influence the development of useful and usable semantic resources.

References

AGLS. *Australian Government Locator Service Document Vocabulary Encoding Scheme*. Canberra: AGLS, Last updated 2013-07-09. Available at <<http://www.agls.gov.au/documents/agls-document/>> Last checked 2015-04-22.

AAT. *Art & Architecture Thesaurus*. Los Angeles: J. Paul Getty Trust. Available at <<http://www.getty.edu/research/tools/vocabularies/aat/>> Last checked 2015-04-22.

APS. American Physical Society. APS Journals. < <http://journals.aps.org/>> Last checked 2015-05-10.

Bowker, Geoffrey C., and Susan Leigh Star. *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press, 1999.

Foskett, A. C. "Misogynists all: A study of critical classification." *Library Resources & Technical Services* 15:117-121 (1971).

Furner, Jonathan. Dewey deracialized: A critical race-theoretic perspective. *Knowledge Organization* 34 (No. 3): 144-168 (2007).

Furner, Jonathan, and Anthony W. Dunbar. "The treatment of topics relating to people of mixed race in bibliographic classification schemes: A critical race-theoretic approach." In *Knowledge Organization and Global Information Society: Proceedings of the Eighth International ISKO Conference*, 13-16 July 2004, London, UK, ed. C. McIlwaine. Würzburg: Ergon. (2004) 115-120.

HHS. *Household Products Database*. Washington, D.C.: U.S. Dept. of Health and Human Services, Last updated 2014-08. Available at <<http://householdproducts.nlm.nih.gov/>> Last checked 2015-05-02.

ISIC. *International Standard Industrial Codes*. New York: U.N. Statistics Division, [2008]. Available at <<http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=27>> Last checked 2015-04-22.

ISO 3166. *Country Codes*. Geneva: International Standards organization. (http://www.iso.org/iso/country_codes.htm) Last checked 2015-05-02.

Google. *The Knowledge Graph*. Sunnyvale, CA: Google. Available at <<https://www.google.com/search/about/insidesearch/features/search/knowledge.html>> Last checked 2015-05-03.

LCNAF. *Library of Congress Name Authority File*. Washington, D.C.: Library of Congress. Available at <<http://id.loc.gov/authorities/names.html>> Last checked 2015-04-22.

LCSH. *Library of Congress Subject Headings*. Washington, D.C.: Library of Congress. Available at <<http://id.loc.gov/authorities/subjects.html>> Last checked 2015-05-02.

NY Times. *The New York Times Topics*. New York: New York Times. Available at <<http://topics.nytimes.com/top/reference/timestopics/>> Last checked 2015-04-22.

NAICS. *North American Industry Classification System*. Washington, D.C.: U.S. Dept. of Commerce. Census Bureau, Last revised 2014-11-05. Available at <<http://www.census.gov/eos/www/naics/>> Last checked 2015-04-22.

Olson, Hope A. *The power to name: Locating the limits of subject representation in libraries*. Dordrecht, the Netherlands: Kluwer Academic, 2002.

Olson, Hope A., and Rose Schlegl. "Standardization, objectivity, and user focus: A metaanalysis of subject access critiques." *Cataloging & Classification Quarterly* 32 (2): 61-80 (2001).

OMB. *Federal Enterprise Architecture Framework: Version 2*. Washington, D.C.: U.S. Office of Management and Budget, 2013. Available at <https://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/fea_v2.pdf> Last checked 2015-05-02.

PACS 2010. *Regular Edition*. Washington, D.C.: AIP Publishing, 2010. Available at <<http://www.aip.org/publishing/pacs/pacs-2010-regular-edition>> Last checked 2015-05-02.

Tennis, Joseph T. "Subject ontogeny: Subject access through time and the dimensionality of classification." In *Challenges in knowledge representation and organization for the 21st century*:

Integration of knowledge across boundaries: Proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain, ed. María J. López-Huertas, 54-59. Würzburg, Germany: Ergon, 2002.

TGN. Getty Thesaurus of Geographic Names. Los Angeles: J. Paul Getty Trust. Available at <http://www.getty.edu/research/tools/vocabularies/tgn/> Last checked 2015-05-10.

ULAN. *Union List of Artist Names Online*. Los Angeles: J. Paul Getty Trust. Available at <http://www.getty.edu/research/tools/vocabularies/ulan/> Last checked 2015-05-03.

UNSPSC. *United Nations Standard Products and Services Code*. 2014. Available at <http://www.unspsc.org/> Last checked 2015-05-02.

USGS. *Geonames Database*. Washington, D.C.: U.S. Geological Survey. Board on Geographic Names, Last modified 2015-03-13. Available at <http://geonames.usgs.gov/> Last checked 2015-05-02.